

ODP-8-2200
7 DEC 1973

MEMORANDUM FOR: Director of Central Reference

FROM : Clifford D. May, Jr.
Director of Data Processing

SUBJECT : Draft of Proposal for a Centralized
Community Bibliographic and Document
Retrieval System Operated by CIA

*fil in
project
folder*

1. Attached for your review and comment is a revision of your memorandum on the same subject which you addressed to me on 19 October. We discovered a few costs which had not been reflected in your proposal and in attempting to introduce this new information we found it advantageous to separate the discussion of costs from the discussion of the bibliographic and retrieval systems themselves. It seems to us that the resource implications of the several options are easier to following using this format. In melding our costs with yours we may have not eliminated overlaps and we suggest a careful review of para. 5. and the attached table of costs. Note too the need for a specific estimate of maintenance costs in para 5.d.

2. Another change has been the introduction of a third option (para 3.c.) for Community access to bibliographic information. As noted, this third option is a mixture of the off-line and on-line options contained in your proposal.

3. Unlike the attachments to the original paper, which added ADSTAR costs to the costs of the bibliographic options, the table attached to this draft covers only the bibliographic options. The costs of expanding ADSTAR are covered in the paragraphs on that system but, because of uncertainty surrounding the costs of expanding ADSTAR, tabular presentation of this information seemed to serve little purpose.

4. Perhaps we can discuss the modified proposal at our monthly meeting now scheduled to be held at 4:00 on 11 December.



Clifford D. May, Jr.

STATINTL

Atts: a/s

Distribution:

Original - Addressee, w/att.

2 - O/D/ODP

1 - ODP Registry

STATINTL

O/D/ODP [redacted] caj/6 December 1978

ODP-8-2184

MEMORANDUM FOR: Chairman, DCI Intelligence Information
Handling Committee

FROM : Clifford D. May, Jr.
CIA Member, IHC

SUBJECT : Proposal for a Centralized Community
Bibliographic and Document Retrieval
System Operated by CIA

1. Proposal: This memorandum proposes that Intelligence Information Handling Committee study the feasibility and desirability of adopting CIA's RECON bibliographic index and ADSTAR micrographic document storage and retrieval system as a Centralized Intelligence Community Bibliographic and Document Retrieval System, managed and operated for the Community by CIA.

2. Background: a. The RECON subject file, from which the proposed Community data base would be derived, has several advantages over other computer-based document indexing systems currently used by NFIB agencies. Initiated in 1968, the RECON file is the largest and most comprehensive subject index to intelligence reports in the Community. As of September 1978 the file contained 3,000,000 index records. RECON offers access to virtually all substantive intelligence documents originated (given general distribution) by the CIA, DoD, DIA, Air Force, Army, Navy, NSA, State, and NPIC, and some documents from other government agencies of the United States [REDACTED]. The data base contains both raw and finished intelligence reports, includes both collateral intelligence and Sensitive Compartmented Information (SCI), and the area coverage is worldwide. Subjects indexed include government, politics, society, culture, science and technology, transportation, communications, business, commerce, industry, finance, commodities (both strategic and non-strategic), products (civilian and military), resources (including labor and military manpower), and the armed forces. In brief, no area of interest to intelligence is overlooked. Open literature, non-CIA cables, and [REDACTED] reporting are included on a selective basis.

STATINTL

STATSPEC

b. The full RECON data base is stored in machine-readable form and is searchable by computer via any one or a combination of the elements used to describe each document. These include the bibliographic description (title, issuing agency, post or origin, date, report number, security classification and dissemination restrictions); area codes (China and the Soviet Union are subdivided to the province and oblast level, respectively); specific place names where appropriate; subject codes; and keywords. The 320 subject codes are standardized broad subdivisions, more than one of which can be assigned to any single document by the indexers in CIA's Office of Central Reference (OCR). The keywords are non-standardized terms added by the indexer based on review of the title and document text; these individual keywords supplement the broader subject codes and thus refine the retrievability of each individual document. The flexibility of such an indexing system allows it to easily accommodate new subject indexing requirements.

c. RECON has an historical depth of 10 years and is the most up-to-date general purpose subject index to intelligence documents available. Approximately 85-90 percent of incoming documents are available for computer search of the index records within eight days after receipt, and by July 1979 this figure will be reduced to three days. Portions of the RECON data base are now available to the Community via COINS, and the total data base itself has been queried on a limited basis by OCR analysts for all NFIB agencies continually since its development. When CIA's earlier bibliographic retrieval system, known as "Intellofax," was in operation, then non-CIA use of the CIA index to intelligence reports was about 45 percent of total queries. With the initiation of the AEGIS/RECON system in 1967-68, however, CIA management placed severe limits on other agency access to these bibliographic records because of substantial reductions imposed on CIA resources. Even under this restriction, however, non-CIA use of the data base has crept upward, and during the first half of CY-1978 the entire data base was queried over 800 times by non-CIA NFIB agencies (approximately 26% of total queries during this period). During the same period, the finished intelligence portion of the RECON data base, which is part of the COINS system, was queried via COINS by non-CIA NFIB agencies over 1,200 times.

d. Bibliographic services must be supplemented by document retrieval capabilities. To ensure speedy and efficient retrieval, CIA is building an Automated Document Storage and Retrieval (ADSTAR) System, which is scheduled to enter operation in November 1979. Designed to operate either in batch or online mode, ADSTAR will store documents on microfilm but digitize these images for transmission over broad-band communications links to remote display terminals and printers.

3. Community Options for Bibliographic Service:

a. Offline Service

(1) The least costly approach of providing RECON bibliographic records to the Community would simply entail offering increased service from the system in its present configuration to other NFIB members. Under this arrangement, a non-CIA analyst presents his research request in writing or over the phone to an OCR area reference analyst, who queries the RECON data base and then mails the printed listing of records to the original requester.

(2) The primary disadvantages of this system are the delays involved in having to mail the request and document listing. The existence of an intermediary (the OCR area reference analyst) between the end user of the data and the data base itself can also be a disadvantage, but not without some positive aspects. Among the disadvantages, the requester may have no way of knowing how large or small a document listing he will be getting until he receives it from the area reference analyst. Any revision of his query to make his request either more inclusive, more selective, or otherwise more appropriate for retrieving precisely what he needs can only be made after the query has been run and the complete document listing is received through the mail. On the positive side, the intermediary reference analyst usually has a better knowledge than the requester of the subject indexing codes and keywords (including how they have been used), and he can often translate the requester's needs into a more effectively worded query than if the requester is left to his own devices.

b. Direct Online Service

(1) If CIA's RECON data base is to be made available to all other NFIB agencies, there is a preferred alternative to merely expanding the operation described above. This would be to provide online access to the data base (stored at CIA Headquarters) via remote visual display terminals (VDTs) in other agencies. Such access could be made available on a 24-hour/day basis if necessary. Bibliographic references displayed on these remote VDTs could be printed immediately on medium-speed (300 lines/minute) printers co-located at each VDT. In this connection it should be pointed out that since the fall of 1973 a variety of intelligence analysts in CIA have been successfully querying the entire RECON data base directly via the SAFE Interim System¹ remote VDTs without OCR intervention. These analysts were formally trained to search the data base and are provided with guidance when necessary.

(2) The principal advantages of this arrangement include the significantly faster availability of the document citations to the analyst, plus the capability for the analyst to work directly with the data base. The latter feature would enable the analyst to determine if the subject codes and keywords he had chosen were producing references to the kinds of documents he needed; he could also see how large his document listing would be and modify his query parameters if necessary. All this could be done before ordering a printout from the system. For standing requests for index searches the capability to query the data base via the batch mode would be retained, rather than requiring the analyst to repeatedly compose his query at a terminal.

(3) If the online arrangement outlined is adopted, existing data communications systems such as the COINS network should be able to handle the transmission of the RECON bibliographic records from CIA Headquarters to requester terminals located at other NFIB agencies.

¹This is the precursor of the ultimate SAFE system, designed to assist in all aspects of intelligence production.

c. Online Service through Intermediaries

(1) Somewhere between options a. and b. above would be a system in which community customers would be linked to OCR's area reference analysts in a network of computer terminals. Queries would be presented telephonically or via the computer terminal, and the results of the analysts' online search could be displayed on the requester's terminal.

(2) The advantages of this blend of services are clear and have to do with effective, real-time communications between the area reference analyst and his customer. Questions about individual bibliographic references can be answered and the document listing tailored to the customer's needs. The refined listing could then be printed at the customer's printer as in option b.

4. Community Options for Document Retrieval Service:

a. Batch Mode

Under this configuration the CIA ADSTAR system would produce copies of documents after receiving requests either in writing or by computer terminal command, depending upon which form of bibliographic service has been adopted. The documents would be mailed to the requester.

b. Direct Online Retrieval

(1) In its most sophisticated configuration, remote ADSTAR terminals located throughout the Intelligence Community would allow non-CIA requesters to query the CIA's central ADSTAR library and display the text and print hard copies of whichever documents the NFIB analyst selected from his RECON listing.

(2) Such an online document retrieval system, however, could not be developed on the basis of existing data communications systems, such as the COINS network. This is because the bandwidth capacity to handle ADSTAR document image transmissions, which consists of approximately four million bytes per page image, is not available

in existing Community networks. The data transmission problem could be eased somewhat by using advanced data compression techniques, but even such a compressed data transmission would require an estimated one million bytes per page image.

5. Costs:

a. Any expansion of RECON services will require a major redesign of the data base. This redesign, to remove Input/Output bottlenecks and to render RECON capable of responding efficiently to larger online system requirements, would cost an estimated \$250,000, plus annual maintenance of \$100,000. These costs are basic and will be incurred if any major increase in the use of RECON is planned, whichever options are adopted.

b. If option 3.a. is adopted, about ten more document indexers and dissemination personnel would be needed to process the additional material expected to be added to the data base, in addition to indexing certain categories of documents in greater depth to satisfy the anticipated specific needs of various agencies. An additional typist would be necessary for the added input to the data base. Two additional camera operators would be needed in OCR's Microform Processing Branch to handle the increased volume of incoming documents to be filmed. Fifteen more area reference analysts would be needed to handle the added volume of requests.² At least two more clerks would be needed to address and package listings for mailing and to prepare document and courier receipts. An additional direct access storage unit would have to be leased in order to store the greater number of document citations in the data base. No additional computer equipment, software, personnel or floor space would be required. These operating expenses would probably approximate \$600,000 per year.

c. If option 3.b. is adopted, and assuming that the COINS network were used, in addition to the costs cited in para. 3.c. above, a large, dedicated host computer would have to be installed, at a cost close to \$4 million. System software would have to be modified to make the computer program "reentrant," an arrangement enabling the central

²It is extremely difficult to accurately estimate the number of index search requests that would be levied on CIA if RECON were made available to the Community without restriction. However, for the purpose of this memo, it is assumed that the current level of requests would increase five-fold. (This figure is largely a guess, based partly on OCR's experience with non-CIA requesters before controls were imposed on their use of the RECON data base.)

processing unit to handle up to 50 online requesters simultaneously. This would entail a one-time payment to a contractor, and would require approximately three man-years of his work and one calendar-year of time. An extra programmer and technician would each be needed in OCR's computer support unit to work with the contractor during the software modification and later to maintain this software and troubleshoot the system's operation. In addition to making the host computer operational for RECON, a number of other tasks would be required. The software interfaces connecting the computer, the message processor, and the COINS network would have to be developed. Certain additional software and hardware changes would be needed to adapt the RECON system to accommodate an increased number of users. Also, some combination of software modifications and human intervention may be required to resolve security release problems. Total cost for this effort would approximate \$500,000.

d. To house the host computer approximately 2,500 square feet of computer-grade floor space would be required, and ten positions would be needed for the personnel to operate the computer in a stand-alone environment that is electrically isolated from CIA's other computer facilities. The annual operating costs would include an additional computer programmer, and a computer technician, plus higher equipment maintenance costs. The total of these operating costs is estimated to be about \$220,000 per year for personnel and _____ for maintenance.

e. In addition to the extra personnel--including indexers and microphotographers--already mentioned, a centralized staff of about three or four people (\$60-80,000/year) would probably be necessary to coordinate new indexing requirements from participating agencies; to train personnel to use the system and to provide on-going guidance once the system enters operation; and to handle trouble calls and transmit questions to appropriate operating personnel.

f. Option 3.c. would avoid the costs related to the installation and operation of a host computer and the attendant software development costs referred to in para. c. above, but the use of computer terminals to deliver bibliographic information would entail careful systems design and probably the acquisition of a number of "smart" terminals for use by OCR's analysts, terminals with the

ability to store information received from RECON and to deliver it on command to the remote customer terminal, which, in this configuration, would not have direct access to the CIA computer housing the RECON data above. Cost figures for such a system cannot be developed without a major study, but the costs should be significantly lower than those associated with the stand-alone host computer.

g. The various costs described above are summarized in the table attached to this memorandum.

h. The costs of Document Retrieval Service Option 4.a. can also be separated into investment and operating expenses. An ADSTAR system augmented to provide Community-wide service would require approximately eight more storage modules to accommodate the assumed 25 percent increase in the number of documents five years old or less that are to be stored in that portion of the system designed to provide immediate retrieval. (These need not be added all at once; two per year could probably take care of the expected annual ADSTAR file growth.) Larger central processing units would be needed to accommodate the greater number of index records and associated support files. For the same reasons more disk packs and disk drives would be needed, the buffer capacity would have to be doubled and at least one other high-speed printer would have to be acquired. If this new centralized document service were to result in a demand for more documents in microfiche, the microfiche output capability would have to be greatly enhanced. Finally, software modifications to the ADSTAR system would be needed. These would all be on-time investment costs, and, while extremely conjectural, would probably total over \$1,000,000.

i. The increased operating costs anticipated for an expanded ADSTAR system would include two additional personnel to intervene in the ADSTAR process to resolve document release questions. Two extra clericals would be needed for packaging, mailing, and preparing document and courier receipts for batch requests for documents. Maintaining the various expanded support files (e.g., MIS and Security Access) would require another full-time employee. For preventive maintenance of the additional equipment, the maintenance contract would cost more. These operating costs would probably come to about \$150,000 per year.

j. Direct Online Retrieval, as in Option 4.b., would require additional outlays for a central processing unit of greater capacity, more software, and (most importantly) the communications system hardware; the latter would include the communication lines themselves as well as the interface equipment, cryptographic systems, and remote access and display stations. Also, as with the online bibliographic retrieval system, appropriate measures would have to be taken to handle security release problems before this system is implemented. We cannot estimate the total of these additional costs without tasking communications specialists to undertake a system study, but undoubtedly the costs would be substantial.

6. Funding: a. Funding could be accomplished in at least four different ways, each of which has its advantages and disadvantages. One possible method involves user agencies supplying personnel to CIA according to a ratio proportionate to the additional input burdens each agency would impose on the RECON system plus the use each agency made of the system. This method has been used between CIA and NSA for reference support under Project Millstream. Its applicability when a number of agencies are concerned, however, is questionable. There is the problem of allocation of manpower compensation from individual agencies whose costs to the system are fractions of many years. There are also the problems attendant with periodic replacement of personnel and with the loss of control by CIA in applying its own personnel selection procedures and standards to all of the people working in the CIA.

b. A second alternative would be to have user agencies transfer funds to the CIA to pay for their portion of the input and use made of the RECON/ADSTAR system basically a "charge-back" system. This would be similar to an arrangement during the 1950's and early 1960's between the State Department and the CIA, whereby the latter transferred funds to the State Department to pay for the CIA's use of State Department biographic files. This approach is easier to arrange and manage than the transfer of personnel, but is complicated by the situation in which a number of agencies must defend a portion of their budgets that are allocated to a program run by another agency. Furthermore, this alternative does not address the question of personnel, so a situation could arise in which the CIA had enough money, but had not been authorized enough additional slots for the people needed to operate the system.

c. A third way would be to have those developing and operating costs of the system that are associated with Community service (including the additional positions required) made part of the budget of the Information Resources Office, RMS and to charge the IRO with defending this portion of its budget each year before Congress. A peculiarity associated with this arrangement would be that the investment and operating funds for an essentially integrated system would have to be split between two budgetary sources, and potential complications could develop if differing budgetary priorities ever arose between the IRO and the CIA.

d. The fourth possible method would be to increase CIA/OCR's budget to allow it to finance the development and operation of the system itself. Such a proposal was made by OCR as an "enhanced" option in its FY-1980 program call, but it was rejected. If adopted, however, it would have the advantage of administrative simplicity and would avoid any complications arising from splitting the source of funds for developing and operating the system among different organizations.

7. Time Required for Implementation: a. Any planned expansion of the CIA's bibliographic and document retrieval system would require a thorough and detailed study of at least six months' duration, plus time to hire whatever additional personnel the study will have called for.

b. Off-line bibliographic service (option 3.a.) could be implemented as soon as additional service personnel were hired, possibly as early as six months after completion of the initial six-month preliminary study, assuming that the requisite floor space could be acquired.

c. The more advanced approach of providing online bibliographic access (option 3.b.) would probably require at least two years after completion of the initial six-month study. During this period, software modifications would have to be accomplished, additional equipment would have to be acquired and installed, and non-CIA agencies would have to program their budgets for the communications equipment and remote terminals they must fund. About the same time would be required to implement a system of online service through Intermediaries using a network of computer terminals (option 3.c.).

d. Centralized document retrieval would be impossible for the CIA until after the ADSTAR system had been implemented and operationally tested for at least six months.

This would make ADSTAR available for Community-wide use no earlier than June 1980, and then only for batch retrieval (option 4.a.).

e. An online ADSTAR system that serviced non-CIA agencies via remote work stations (option 4.b.) would take at least two more years for programming user-agency budgets, and acquiring and installing the necessary additional equipment. FY 1982 would be a conservative target date.

8. Recommendation: a. We recommend that the IHC sponsor a study in depth of the Community's bibliographic and document retrieval needs to determine whether centralized services of the kinds described above would serve the Community's interests. The study should emphasize user requirements, system architecture (including communications), and precise investment and operating costs, together with offsetting savings to be made by reducing on-going activities or planned new ventures for which substantial expenditures are planned. Other aspects of the proposal which need research are the security restrictions to be imposed, and floor space requirements for machines and people.

b. If this study demonstrates that centralized services are desirable and economical, we recommend the adoption of RECON and ADSTAR in whichever of the configurations described above most effectively meets the needs of the Community, provided a suitable answer can be found to the questions of manning and funding the Community support.

Clifford D. May, Jr.

Att: a/s

Approved For Release 2002/01/08 : CIA-RDP84-00933R000100300003-7

Requirement	Option 3.a.			Option 3.b.			Option 3.c.		
	Positions	One-Time	Recurring	Positions	One-Time	Recurring	Positions	One-Time	Recurring
Redesign RECON		250,000	100,000		250,000	100,000		250,000	100,000
<u>Bibliographic Service</u>									
<u>Off-line</u>									
- 13 Index/Dissem/Clerical, 2 Camera Op., 15 Area Reference Analysts	30		600,000	30		600,000	30		600,000
(Add. Direct Access Storage Unit			?			?			?
<u>On-line (Direct)</u>									
- Host Computer					4,000,000*				
- Software					500,000				
- 10 Operators, 1 Tech, 1 Systems Analyst, 3 Requirements Coord.				15		280,000			
- Operating Costs						?			
<u>On-line (Intermediary)</u>									
- Smart Terminals								250,000(?)	
- Software								250,000(?)	
Sub-Totals	30	250,000	700,000(?)	45	4,750,000*	980,000(?)	30	750,000(?)	700,000(?)
Total Annual Cost			50,000 **			950,000 **			**150,000
Assuming 5-Year System Life			750,000(?)			\$1,930,000(?)			\$850,000

*Plus 2500 sq. ft. of floor space.

**Annual figures represent 1/5 of the one-time totals shown in preceding column.

Approved For Release 2002/01/08 : CIA-RDP84-00933R000100300003-7